OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: February 27, 2003, 03:34:07 ; Search time 1693 Seconds
(without alignments)
6189.297 Million cell updates/sec

Title: US-09-677-653A-50
Perfect score: 3374
Sequence: 1 MGDAGVASQRPHNRGTRNV.........GKIAARVRARRARRRAARAN 647

Scoring table: BLOSUM62

Xgapop	10.0 ;	Xgapext	0.5
Ygapop	10.0 ;	Ygapext	0.5
Fgapop	6.0 ;	Fgapext	7.0
Delop	6.0 ;	Delext	7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+_p2n.model
-O=/cgn2_1/USPTO_spool/US09677653/runat_20022003_164315_1262/app_query.fasta_1.839
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US09677653 @CGN 1_1_1132 @runat_20022003_164315_1262 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*

1:	em_estba:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_htc:*
9:	em_estl:*
10:	gb_est2:*
11:	gb_htc:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_ests:*
15:	em_estfun:*
16:	em_estom:*
17:	gb_gss:*
18:	em_gss_hum:*
19:	em_gss_inv:*
20:	em_gss_pln:*
21:	em_gss_vrt:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	121.5	3.6	3775	11	BC028165	BC028165 Homo sapi
2	116	3.4	1535	13	BM552027	BM552027 AGENCOURT
3	112	3.3	900	12	BF167575	BF167575 60177428
4	111.5	3.3	1069	14	BM913440	BM913440 AGENCOURT
5	111.5	3.3	1269	14	BQ712831	BQ712831 AGENCOURT
6	109.5	3.2	2128	13	BI082474	BI082474 60287781
7	109	3.2	3572	11	BC010575	BC010575 Homo sapi
8	108.5	3.2	915	13	BI413164	BI413164 60299903
9	108.5	3.2	1776	14	BM926649	BM926649 AGENCOURT
10	108	3.2	861	12	BG168931	BG168931 60232028
11	107.5	3.2	869	9	AL578091	AL578091 AL578091
12	106.5	3.2	686	17	AG139016	AG139016 Pan trogl
13	106.5	3.2	820	13	BI650823	BI650823 60297718
14	106	3.1	959	13	CNS03M9N	CNS03M9N Tetraodon
15	106	3.1	983	14	BQ649851	BQ649851 AGENCOURT
16	106	3.1	1060	13	BM549598	BM549598 AGENCOURT
17	106	3.1	1061	17	CNS03379	CNS03379 Tetraodon
18	105.5	3.1	635	9	AI541596	AI541596 S002772.5
19	105.5	3.1	783	13	BI652061	BI652061 60298842
20	105.5	3.1	1090	14	BM479488	BM479488 AGENCOURT
21	105.5	3.1	1294	13	BI409855	BI409855 60296197
22	105	3.1	695	13	BJ244281	BJ244281 BJ244281
23	105	3.1	1087	12	BF973817	BF973817 60224191
24	105	3.1	1485	14	BM907789	BM907789 AGENCOURT
25	104.5	3.1	908	12	BE736099	BE736099 60130584
26	104.5	3.1	962	14	BQ955611	BQ955611 AGENCOURT
27	104	3.1	891	14	BQ670497	BQ670497 AGENCOURT
28	104	3.1	973	12	BF981727	BF981727 60236079
29	104	3.1	978	17	BQ714511	BQ714511 AGENCOURT
30	104	3.1	2391	17	BH771003	BH771003 LMGtag72
31	103.5	3.1	679	13	BM406277	BM406277 EST580604
32	103.5	3.1	830	14	BM985604	BM985604 2_A12_T3
33	103.5	3.1	963	14	BQ899717	BQ899717 AGENCOURT
34	103.5	3.1	1067	12	BG168781	BG168781 60230073
35	103.5	3.1	1092	14	BQ676763	BQ676763 AGENCOURT
36	103.5	3.1	1858	13	BM557460	BM557460 AGENCOURT
37	103	3.1	937	13	BM461960	BM461960 AGENCOURT
38	103	3.0	1031	17	CNS05IRX	CNS05IRX Tetraodon
39	102.5	3.0	663	14	BQ917203	BQ917203 QHB20J22.
40	102.5	3.0	750	13	BI768562	BI768562 603057176
41	102.5	3.0	942	14	BQ930180	BQ930180 AGENCOURT
42	102.5	3.0	1060	13	BM424198	BM424198 AGENCOURT
43	102.5	3.0	1287	13	BM546426	BM546426 AGENCOURT
44	102	3.0	1094	17	AL255159	AL255159 Tetraodon
45	102	3.0	1259	14	BQ960839	BQ960839 AGENCOURT

ALIGNMENTS

RESULT 1
BC028165

LOCUS   BC028165   3775 bp   mRNA   linear   HTC 01-MAY-2002
DEFINITION Homo sapiens, protocadherin 17, clone IMAGE:5245587, mRNA.
ACCESSION   BC028165
VERSION   BC028165.1 GI:20380902
KEYWORDS   HTC.
SOURCE   Homo sapiens.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3775)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-APR-2002) National Institutes of Health, Mammalian

```
REMARK          Gene Collection (MGC), Cancer Genomics Office, National Cancer
                Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
                USA
COMMENT         NIH-MGC Project URL: http://mgc.nci.nih.gov
                Contact: MGC help desk
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: Life Technologies, Inc.
                cDNA Library Preparation: Life Technologies, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: National Institutes of Health Intramural
                Sequencing Center (NISC),
                Gaithersburg, Maryland;
                Web site: http://www.nisc.nih.gov/
                Contact: nisc_mgc@nhgri.nih.gov
                Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
                Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
                Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
                Hansen,N., Ho,S.-L., Karlins,E., Laric,P., Legaspi,R., Maduro,Q.L.,
                Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
                Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C.,
                Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A.,
                Zhang,L.-H. and Green,E.D.
                Clone distribution: MGC clone distribution information can be found
                through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
                Series: IRAK Plate: 62 Row: n Column: 5
                This clone was selected for full length sequencing because it
                passed the following selection criteria: matched mRNA gi: 14589926
                This clone has the following problem: frame shifted.
FEATURES        Location/Qualifiers
source          1..3775
                /organism="Homo sapiens"
                /db_xref="LocusID:27253"
                /db_xref="taxon:9606"
                /clone="IMAGE:5245587"
                /tissue_type="Brain, fetal, whole pooled"
                /clone_lib="NIH_MGC_121"
                /lab_host="DH10B"
                /note="Vector: pCMV-SPORT6"
BASE COUNT       916 a   1117 c   1031 g   711 t
ORIGIN          1117 c
DB:
Query Match:     11
Best Local Similarity:    3.60%
Percent Similarity:      22.16%
Score:                  121.50
Pred. No.:               0.648
Alignment Scores:

US-09-677-653A-50 (1-647) x BC028165 (1-3775)
```

```
Qy     12  HisAsnArgArgGlyThrArgAsnValArgValSerAlaAsn-ThrValThrValAsnGl  31
           ||||                    || |||| ||||| |||:::::
Db    318  CACAATGCCAAGTGCCAAGCTGTCCCTGCGAGTGTTCGCCAACGACA------------   363

Qy     31  yArgArgAsnGln------ArgArgArgThrGlyArgGlnValSerProProAs  47
           ||||||||                   || |||||| ||||||
Db    364  -AGGAGAATCTGCATGATCAAGGTAGGAGATCCAGGACATCAACGACAACGGCCGCCTCCTTC  422

Qy     47  pAsnPheThrAlaAlaAlaGlnAspLeuAlaLeuSerLeuAspAlaAsnThrValThrProH  67
           |||                || |||  ||| || |||||||||||
Db    423  TCCTCGGACCAGATCGAA-ATGGACAATCTCGGAGAAGCTGCTCCCGGGCACC--CGCTT   478

Qy     67  eProAlaAsnLeuSerSerMetProGluPheArgAsnTrpAlaLysGlyLysIleAspPle  87
           |||||          :::||
Db    479  CCCCCTCACGCGGCACACATGACCCC----------------------------GACGC   508

Qy     87  uAspSerAspSerIleGlyTrpIleTyrTyrLeuAspProAlaGlyAlaThrPh  106
           :::            |||          |||
Db    509  CGGCGAGAATGGGCTCGCACCTACCTGCCACTCACGGCGCTCTTCTTTGGGACT  568

Qy    107  -SerAlaArgAlaValGlyGluTyrSerLysIleProAspGlyLeuValLysPheSerVa  126
           ||                      :::|
Db    569  GGAGGTTAAGCTCCCGCGCCGGGACGCGCACCAACAGTTCCCAGAACTGGTCATCCGAACGGCTCT  628

Qy    126  IAspAlaGluIleArgGluValTyrAsnGluGlyCysProValValThrAspValSerVa  146
           ||||| ||||                ||     |||||:|||
Db    629  GGAC--------------CGCGGAGCAACAGAATCACCATACGCTCGTGCTGACT------------   668

Qy    146  lProLeuAspGlyArgGlnThrTrpSerProLeuSerPheSerSerPheProMetPheArgThrAl  166
           ||||||||||              :::           |||
Db    669  -GCCCTGACGGTGCGCAG------------------------CCTCCACGTTCCGCACCGT  706

Qy    166  aTyrValAlaValAlaAsnValGluAsnLysGlyGluMetSer------LeuAspValAla  184
           :::                            |||
Db    707  ACAGATCAACGTGAAGGTGATTGACTCCAACGACAACAACACCCGGTCTTCGAGGCGCCATC  766

Qy    184  nAspLeuIleGluTrpLeuAsnAsnLeuAlaAspTrpArgTyrTyrValValAspSerGluG1  204
           ||:::||            ||        |||||||
Db    767  CTACTTGGTGGAACTGCCCCGAGAACGGCTCCGCTGGGTACAGTGGTCATCGAT--------   818

Qy    204  nTrpIleAsnPheThrAsn--------AspThrThrTyr  214
           |||                ||  |||
Db    819  -CTGAACGCCACCGACGCCCGATGAAGGTCCCAAATGGTGAAGTGCCTGCTACTCTTTCAG  874

Qy    215  ---TyrValArgIleLeuArgValLeuArgProThrThrArgValProAsnProThrGluGl  233
           ||||||                    |||
Db    875  CAGCTACGGTGCCTGACCCGGTG---CGGGAGCGCTCTTCTCCATC---GACCCCAAGACCGG  928

Qy    233  yLeuValArgThrValSer------AspTyr----  241
           ||||||                ||||
Db    929  CCTAATCCGTGTGAGAGGGCCAATCTGGACTMATGAGGAAAAACGGGGATGCTGGAGATTGACGT  988

Qy    242  ------------------------ArgLeuThrTyrLysAl  247
           |||
Db    989  GCAGGCCCGAGACCTGGGCCTTAACCCTATCCCAGGCCCACTGCAAAGTCACGGTCAACT  1048

Qy    247  aIleThrCysGluAlaAsnMetProThrLeuValAlaAspGlnGlyPheThrpIleGlyGlyGl  267
           |||                |||
Db   1049  CATCGACCGCAACGACAATGCCGCCTCCATC------GGTTTC------------------  1085

Qy    267  nTyrAlaLeuThrProThrSerLeuProGlnThrTyrAspValSerGluAlaTyrAlaLeuHi  287
           |||:::|                    ||||||||||
Db   1086  ----------------GTCTCGCTGCGCCGCCAGGGGGCGCTGAGCGAGGCC----------  1118

Qy    287  sThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeuAlaPheValTrp--------  304
           ||| |||                |||
Db   1119  ------------GCCCTCCCGGCACCGTCATCGCGTCGGGGTCACTGACCGGA-------  1165

Qy    305  -AlaGly------------LeuProGlnGlyGlyThrAlaProAl  315
           |||||                          |||||||
Db   1166  CTCTGGCAAGAACGGACGACAGCTCGAGTGTCGGGTCCTAGGCGAGGAGGACGGGCGGCGG  1225

Qy    315  aGlyThrProAlaTrpGluGlnAlaSerSerGlyGlyTyrLeuThrArgArgHisAsnGl  335
           |||
Db   1226  CGGG------------------------------------------------------GG  1231

Qy    335  yThrThrThrPheProAlaGlySerValSerTyrValLeuProGlyLeuGlyPheAlaLeuGluAr  355
           ||| |||                |||
Db   1232  CCTGGGCGGCGGGGCCGGGGTTCGGTCGCCCTTCAGCCTGGAGGAAACTAC-------  1280

Qy    355  gTyrArgAspProAsnAspGlySerTrpThrAspPheAlaSerArgAlaGlyAspThrValThrPh  375
           |||:::|                    |||||||
Db   1281  --------------------GACAACTTCTACACG----------------------------  1306

Qy    375  eArgGlnVal------AlaValAspGluValValValThrAsnAsnProAlaGlyGl  392
           |||            ||||||
Db   1307  CCGGCGCCGCTGGACCGGACACGACAGGAGCAGCAGACGTACACACCGGGCGCGGGGGACGG  1366

Qy    392  yGlySerAlaPro--------ThrPheThrValArgVal  402
           ||||||            ||:::||||
Db   1367  GGGCTCTCCTCCCTCAACTCCACCAAGTCGTTCGCGATCAAGATTTCTAGACGAGAACGA  1426

Qy    403  ----ProPro-SerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeuGluThrA  421
           |||||         ::|
Db   1427
```

```
Db  1427 CAACCCGCTCGGTTCACCAAAGGGTCGTACGTTGCTTCAGGTGGTGCACGAGAACATCCC 1486
Qy   421 rgProSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrValA 441
          ||||:::|      ||| |||    |||  |||||| ||||||
Db  1487 GGGAGAGTACCNGGCNTCNGTG-CTCGCCCAGGATCCC---GACCTGGGCCAG------ 1535
Qy   441 laAsnAsnProLysIleGluGlnSerLeuLeuLeuLysGluThrLeuGlyCysTyrLeuValH 461
          :::|||                  :::||||
Db  1536 ------AACGGCACGTATCCTACTCTATCCTGCCCTCCACATCGGCACGTGTCTATCT 1590
Qy   461 isSerLysMetArg---AsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaV 480
          ||||||          |||         :::
Db  1591 ACACCTATGTTGTCTGTGAATCC---ACGAACGGGGCCANTCTACGCCCTGC 1638
Qy   480 alSerPheAsnAsnProGlyThrTyrGluArgThrArgAspLeuProAspTyrThrGlyIleA 500
          |||||||||||||  :::   ||
Db  1639 GCTCCTTAAC------TTCGAGCAGACCAAGGCTTTGAGTTCAAGGTGCTTGCTA 1689
Qy   500 rgAspSer------------PheAspGlnAsnMetSerThrAlaValAlaHisPheA 515
          ::|||||||                 ||
Db  1690 AGGACTCGGGGGGCGCCCCGCACTTGGGAGAGCAACGCCACGGTGAGGGTGACAGTGCTAG 1749
Qy   515 rgSerLeuSerHisSerCysSerSerIleValThrLysThrTyrGlnGlyTrpGluGly---- 533
          |||             :::||||          ||| |||
Db  1750 ACGTGAATGACAACGCGCCAGTGAATCGTGCTCCCCACCGTGCAGAACAACACACCGCGGAGC 1809
Qy   533 ----ValThrAsnValAsnThrProPheG 542
                     :::   ||
Db  1810 TGGCAGGTGCCGCCCAACCCTGGCCTGGGCTANCTGTGCCGGCCCCTAGACA 1869
Qy   542 lyGlnPheAlaHisAlaGlyLeuLeuLys--------------AsnGluGluLeuL 556
          ||             :::|||||
Db  1870 GCGGACTTCGGCGGAGAAGCGGGCGTCTCACCTACGACATCGTGGGACGACAGCACCAC 1929
Qy   556 euCysLeuLeuAspAspLeuAlaThrArgLeuThrGlyValTyrProAlaThrArgAspAsn- 575
          ||                    :::|||
Db  1930 TGTTTGAGATCGAACCCGTTCCACGCGCCGAGAGATCCGCACACCCTTTCTGGGAGACG 1989
Qy   575 ----------PheAlaAlaAlaValSerAlaAlaPheAlaAlaAsnMetLeuSers 590
                                    ||||||
Db  1990 TGACGCCCGTGGGTGGAGCTGGTGGTGGAAGGTGACGACGAAGGCTACCCTGTCCG 2049
Qy   590 erValLeuLysSerGluAlaThrSerSerIleIleLysSerValGlyGluThrAlaValG 610
          :::|||               ||||||
Db  2050 CAGTGGGCCAAG---------------------CTCATCATCCGGCTCGGTGAGC 2082
Qy   610 lyAlaLeuAlaGlnSerGlyLeuAlaLysLeuProGly 621
          ||:::               ||| |||
Db  2083 GATCCCTTCCCGAGGGGGTACCACGGGTGAAATGGC 2117
```

```
FEATURES             Location/Qualifiers
     source          1..1535
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5549464"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Cloned unidirectionally. Primer: Oligo dT.
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Site_1: NotI;
                     Vector: pCMV-SPORT6; Site_1: NotI;
                     Vector: pCMV-SPORT6; Primer: Oligo dT."
BASE COUNT    220 a    596 c    329 g    385 t    5 others
ORIGIN
```

```

**RESULT 2**
**BM552027**

```
Alignment Scores:
Pred. No.:
Score: 0.531
Percent Similarity: 116.00
Best Local Similarity: 30.28%
Query Match: 20.00%
DB: 3.44%
                        13

                        Length: 1535
                        Matches: 72
                        Conservative: 37
                        Mismatches: 114
                        Indels: 137
                        Gaps: 15
```

US-09-677-653A-50 (1-647) x BM552027 (1-1535)

```
Qy   265 GlyGlyGlyGlnTyrAlaAlaLeuThrProThrSerLeuProGlnThrTyrAspPheValSerGluAlaTyr 284
          ||||||||::::                              |||:::
Db    45 GGAGGTGAAAAGCTGGTGTGAGCCCAGTGGGTTCCCACACTCACCCAGGAGCGTGGCTGG 104
Qy   285 AlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeuAlaPheValTrp 304
          ||                           ||:::|||||||:::
Db   105 GCC-------------AGGACCGGGAGAGGAGCACTGCCCTGCCTGGCCCTGCTC--- 149
Qy   305 AlaGlyLeuProGlnGlyTyrGlyThrArgAlaProAlaGlyGly---ThrProAla---TrpGluGln 322
          ||||||||||  |||              :::|||        :::
Db   150 ------CTTCCGCAGTTAGGGGTGGACCGAGCCTCGCTTTCCCCACTGTTTCTGGAGGGAA 203
Qy   323 AlaSerSerGlyGly---TyrLeuThrTrpArgHisAsnGlyThrThrProPheProAlaGlySe 342
          ||||||          |||                ||
Db   204 GGGGAAGGAGGGGGGTCTTC------------------------------------ 222
Qy   342 rValSerTyrValLeuProGlnGlyPheAlaLeuGluLeuArgTyrAspProAsnAspGlySe 362
          ||:::                              |||:::|||
Db   223 ------------------------------AGGCTGGAGCCAGGCTGGGGGTG 245
Qy   362 rTrpThrAspPheAlaSerArgAlaGlyAspThrValThrPhe------------- 375
          |||  :::                              |||
Db   246 CTGGGTGGGAGAGATGAAGATTTAGGGGGCGGCGCCCTCCTGTTTTTCCGGGGGGGCCACC 305
Qy   376 -ArgGlnValAlaValAlaAspGluValValThrAsnAsnProAlaGlyGlySerAl 395
          ||||||| ||||||                :::|||||||||||||||||
Db   306 CCGGCAAACCCCCCAACCCGCACCCTCTTCCCCCCCGAAATCCGCGGGGGGGGC 365
Qy   395 aProThrPheThrValArgValPro---ProSer--------AsnAlaThrTyrThr----- 409
          |||:::                |||                :::|||||||
Db   366 TCCCCTCCCCTTCCGCTTTCCAATCCTCATCCCCAAATTGTGACTCCGCGCCCTTCACATTACA 425
Qy   409 -------AsnThrValPheArgAsnThrLeuLeuGluThrArgProSerSerArgArgLe 427
          ||             ||||||            |||||||
Db   426 CGCCANGACCCCCTTGCTCTGCCAACCTCACCCAATTCCTCTCCCCTCCCTACCCCTGCA 485
Qy   427 uGluLeuPro---------- 430
          ||                 |||
Db   486 TTTTCTCCGCGGCGCCTGGCTCACACCTGCCTCCCTCCGTCGCCCTCGCCCGGCCCC 545
Qy   546 CTCTGGCCCCTTACGTACGGCCTTCTTCTTGTTCTCTGCTTTTANGCCCCCCCCCCCCCCCGCGGCGTGAGGAGTC 605
          |||:::
```

```
Qy  431 -----------------------MetProProAlaAspPheGlyGlnThrValAlaAsnAs 443
                                   :::
Db  606 GGAAGTGTTTGNNCGGTTTNCGTTNCCCCCGGCCGGGCCGCGGGGCACCGGGCCGCGCGGGCCATCACGTAT 665

Qy  443 nProLysIleGlyGlnSerLeuLeuLysGluThrLeuLeu------------------- 455
        ||||||:::::|||                            |||
Db  666 TCCCCGGCTACGACGATGCACAACTCCTCCGCGTACTCTTACAGTGACGCCATCCGCCGCCCCGCC 725

Qy  456 ------------GlyCysTyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThr 472
                        :::|||
Db  726 ACACATAATCACCCGCTTATTCTCCTATACTCCTTTCTATCCCANCCC----------------- 771

Qy  472 rProAlaSerSerPheGlyAlaValSerPheAsnAsnProGlyTyrGluArgThrArgAs 492
        ||| |||       ||| |||           :::::     |||      |||
Db  772 -CCAACGCCATCTCATAGACATTGTATATACTCCGCAATCGGC--GAGCGTANGCGCCA 827

Qy  492 pLeu----------------------------------------------ProAs 495
        |||                                              |||
Db  828 GCTACGACCACACGCATTTACTTCCCACTCTCCACTCTCTCATATTCTCTGTGTGTCCTCCTC 887

Qy  495 pTyrThrGlyIleArgAspSerPheAspGlnAsnMetSer----------------- 508
        |||  |||   |||     ||||| ::::|||
Db  888 TTACTACGCACAGTTACGCCCCATCCTATCCCTCGATCCGTCCGCCCCTCCGCCACGACAGCCTCC 947

Qy  509 ----ThrAlaValAlaHisPheArgSerLeuSerHisSerCysSerIleValThrLys 526
             :::|||     ||||||||||||:::::|||
Db  948 TGTCCTGTCCCAGTCTCGCTACACTCTTACTCGCCTCCTTTTCTTGTTGCTTCTTACTTCGCCCGC 1005
```

RESULT 3

LOCUS   BF167575/c   900 bp   mRNA   linear   EST 30-OCT-2000
DEFINITION   601774283F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3994773 5',
             mRNA sequence.
ACCESSION   BF167575
VERSION   BF167575.1 GI:11047927
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 900)
AUTHORS   NIH-MGC http://mgc.ncl.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM9213 row: f column: 22
             High quality sequence stop: 700.
FEATURES   Location/Qualifiers
   source   1..900
             /organism="Mus musculus"
             /strain="C2ECH II"
             /db_xref="taxon:10090"
             /clone="IMAGE:3994773"
             /clone_lib="NCI_CGAP_Lu29"
             /tissue_type="spontaneous tumor, metastatic to mammary.
             /stem cell origin."
             /lab_host="DH10B"
             /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
             Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
             Library constructed by Life Technologies. Investigator
             providing samples: Gilbert Smith, NIH"
BASE COUNT   222 a   235 c   284 g   159 t
ORIGIN

Alignment Scores:

Pred. No.:
Score: 112.00   Length: 900
Percent Similarity: 33.33%   Matches: 66
Best Local Similarity: 24.18%   Conservative: 25
Query Match: 3.328   Mismatches: 84
DB: 12   Indel: 98
                                      Gaps: 14

US-09-677-653a-50 (1-647) x BF167575 (1-900)

```
Qy  225 TyrAspValProAspProThrGluGlyLeuValArgThrValSerAspTyrArgGlnThr 244
        |||             :::||||||||:::|||
Db  814 TACTTGTTGTGCCGATCCCGTT-------------------------------------- 794

Qy  245 TyrLysAlaIleThrCysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrp--- 263
        |||:::||| |||::::|||   |||   :::|||||::::|||
Db  793 --GTTGCCCGCGCTTTTATGGCCAGGGCGCGCCCTCAGCTGTGGAATAAGGGCTTTTTCGGA 737

Qy  264 ---IleGlyGlyThrTyrAlaLeu---------------------------ThrPro--- 272
        |||           |||                                  |||
Db  736 GCCATTGGTGACTGGGGGGGGTCTTCGGGGATTAACCTAGCCGGCCATTTTGATGG 677

Qy  273 ----------------ThrSerLeuProGlnTyrAspValSerGluAlaTyr 284
                        :::|||||||
Db  676 AAGAGCCCAGCAGTGCAGTTTGAAAGCTCTGTACCCAGGGCCCCGGCCACAGCAGCT---- 620

Qy  285 AlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeuAlaPheValTrp 304
                                                              :::
Db  619 --GTGCTCAACTCTGGGAGACCT-------------------------------------- 581

Qy  305 AlaGlyLeuProGlnGlnLeuGlyLeuThrAlaProAlaGlyLeuThrProGluGlnLeuAlaSer 324
        |||||||::::|||
Db  580 GCTACAATCCTAGCCGGGAGGAGGTTCTGCT------------------------------ 530

Qy  325 SerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPheProAlaLeuGlySerValSer 344
                  |||:::||| :::|||||||||
Db  529 TCAGCTGCC---AGCCCCTGG-----AATGGAAGTACAnTTCTCTCTCGCCAGGGGGGCG 479

Qy  345 TyrValLeuProGluGlyPheAlaLeuGlyArgTyrAspProAsnAspGlySerTrpThr 364
        :::|||||| |||
Db  478 TGGGTGTTA--------------CTGGTGCTCTGGGAGCCCGAGGGGAC---------- 446

Qy  365 AspPheAlaSerArgAlaGlyAspThrHisValThrPheAsnArgGlnValAlaAlaAlaAlaAspGluValVal 384
Db  446 -------------------------------------------------------------- 446

Qy  385 ValThrArgAsnAsnProAlaGlyGlyGlySerAlaAlaProThrProPheThrrValArgValProPro 404
        |||                             :::||||||||
Db  445 ------CGAGGCCCAGAGAGTGTGCCCACCTTCATA------------------------CCG 413

Qy  405 SerAsnAlaLaTyrArgAsnAlaThrValPheArgAsnThrLeuLeuGluThrArgProSerSer 424
        ||||||:::|||              :::|||
Db  413 --------GCTCAGGAGAGGTCTGGTAACTTACCTGACTCACTCGCCTTCA 269

Qy  425 ArgArgLeuGluLeuProMetProAlaAspPheGlyGlnThrValAlaAlaAsnAsnPro 444
        |||||||:::|||||
Db  356 GACCCTTGTGTTGCCCTTGAAGTCTTTCAGCCTACAGTGTCTGCCCCCCCCCCCCCGCA--- 356

Qy  445 LysIleGluGlnSerLeuLeuLysGluThrLeuGlyCysTyrLeuValHisSerLysMet 464
                  |||:::|||
Db  311 -----------ATCCCCTCCCGCCCCAAGCCTCCACCGAGCCTANCCTCTAACACA-- 311

Qy  465 ArgAsnProValPheGlnLeuThrProAlaSerSerPhe 477
        :::|||         |||:::
Db  268 GATGAACCCCTCCTTCAGTGGCGTCCCTGTGTTCCTCATGG 230
```

RESULT 4

BM913440

LOCUS   BM913440   1069 bp   mRNA   linear   EST 12-MAR-2002
DEFINITION   AGENCOURT_6610241 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5475556
             5', mRNA sequence.
ACCESSION   BM913440
VERSION   BM913440.1 GI:19363819
KEYWORDS   EST.

```
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1069)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DPP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1990 row: j column: 05
             High quality sequence stop: 734.
FEATURES             Location/Qualifiers
     source          1..1069
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5475556"
                     /clone_lib="NIH_MGC_41"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT       214 a    321 c    293 g    223 t    18 others
ORIGIN
```

```
US-09-677-653a-50 (1-647) x BM913440 (1-1069)

Alignment Scores:
Pred. No.:          0.859
Score:             111.50         Length:          1069
Percent Similarity: 32.51%        Matches:          82
Percent Similarity: 22.40%        Conservative:     37
Best Local Similarity: 3.30%      Mismatches:       115
Query Match:        14            Indels:           133
DB:                               Gaps:             17
```

```
Qy     86 AspLeuAspSerAspSerIleGlyTrpTyrTyrPheLysTyrLeuAspProAlaGlyAlaThr 105
          ::||::|:|              |||||||||             ||||||||||||
Db     87 AACCTGGAGAAACCCTTCAGGCGGCTGGTACGACCCGACCTGAGCCGCGGCCACGAG 146

Qy    106 GluSerAlaArgAlaValGlyGluTyrSerLysIleProAspGlyLeuValLysPheSer 125
          |||||::|   |||              :::|||                ::||||
Db    147 GAGGCGGAAGGCGCGGC----GGGCAG----GCCCTACGAGATGCTGGCTATGAGTTGAC 197

Qy    126 ValAspAlaGluIleArgGluIleTyrAsnGluCysProValValThrAspValSer 145
          |||||               :::|      ||||         |||||||
Db    198 ATC------------------------TGCTTCACCTCAGTGCAG 218

Qy    146 ValProLeuAspGlyArgGlnTrpSerLeuSerIlePheSerProMetPheArgThr 165
          |||||::|               ||||||              ||||||

Qy    166 AlaThrTyrValAlaValAlaAsnValGluAsnLysGluMetSerLeuAspValValAsnAsp 185
          |||||                 ||||||           :::|||
Db    167 AlaThrTyrValAlaValAlaAsnValGluAsnLysGluMetSerLeuAspValValAsnAsp 185

Qy    186 LeuIleGluTrpLeuAsnAsnLeuAlaAspThrArgTyrValValAlaAspSerGluGlnTrp 205
          |||||||||          ::|          ||||||
Db    267 ATG-------TGGCTGCCAGTGGTGAGGACTTGGGCGC-------- 296

Qy    206 IleAsnPheThrAsnAspThrThrTyr--------------- 214
```

```
Db    297 ---------------CTCAAGACGCGGCACTATTGGGGGTGGGTGTCTAACCGGTCTCAATAAGCAGAAACT 347
             |||::|
Qy    215 -----------TyrValArgIleArgValLeuArgProThrTyrAspValProAsp 229
                         ||||::::||::|  |||||||||||||
Db    348 GCTGCAAAGCACATGGTGAGGCCCAGGTGAAGATCTGGAGGCGGCCTCCTTATGAGTCCCACA 407

Qy    230 ProThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysSerAlaIleThr 249
             |||
Db    408 CCT------------------------ 410

Qy    250 CysGluAlaAsnMetProThrLeuValAspGlnGlyGlyPheArgProIleGly------- 265
             |||         |||                      ||||||
Db    411 -----------CCGATGGAGCCCAGACCCATCCTTTCTACAGCAACATCAGTAAGGAT 455

Qy    266 GlyGlnTyrAla---LeuThrProThrSerLeuProGlnTyrAspValSerGluAlaTyr 284
             :::||||||||   ||||||           ||||||                ||
Db    456 CGCAGGTTATGCAGACCCTCACAGAAGGAGTTACCCTCGTGGAGAG-TCTGAAGGATAC 514

Qy    285 AlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaAlaLeuAlaPheValTrp 304
             ||||||                            |||
Db    515 ---------------------CAGAGCTCTGCCCTTCTGGAATGAAGAAAT 550

Qy    305 AlaGlyLeuProGlnGlyGlyThrAlaProAlaGlyThrProAlaTrpProGluGln----- 322
             |||||||              |||||||             |||||
Db    551 AGTTCCCCAGATCAAGGAGGGGAAACGTGTACTGATTGCAGCCCATAGGCAACAGCCTCG 610

Qy    323 -------AlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPhePro 339
                    :::||||||           |||                ::||:|||
Db    611 GGGCATTGTCAAGCAATCTGGAGGTGTCTCTGAAGAGGCTATCATGGAGCTGAA----CCT 667

Qy    340 AlaGlyLysSerValSerTyrValLeuProGluGlyPheAlaLeuGluArgTyrAspProAsn 359
             ||||||       |||                ||||||              :::
Db    668 GCCGACTGGGTATTCCCATTGTCTATGAATTGGACAAGAAACTTGAAGCCTATCAAGCCCAT 727

Qy    360 AspGlySerTrpThrAspPheAlaSerAlaGlyAspThrValThrProHisArgGlnValAla 379
             |||                            |||
Db    728 GCCAGTTTCTGG-----------GGGGATAGAAGACGGTGCGCCAA---GCC 766

Qy    380 ValAspGlyValValThrAsnAsnProAlaGlyGlyGly------------------ 393
             |||                  ||||||
Db    767 ATGGAAGCTGTGGCTCGCCAGGCAAGGCCCANAAGTGGAAGCGGGGGGGGAAGAGGATACTGT 826

Qy    394 --------SerAlaProThrPheThr-ValArg---ValProProSerAsnAl 407
                     |||         |||           |||||
Db    827 CCCCAGGAGGACACCCTCCCCCGTGTTTGCCCCTCGGCCCCCGCCCCCACCTGGGAAATGC 886

Qy    407 aTyrThrAsnThrVal 412
             |:|  |||  :::
Db    887 CTTCCCTGAACCCCATC 902
```

```
RESULT 5
BQ712831
LOCUS        BQ712831                1269 bp    mRNA    linear    EST 16-JUL-2002
DEFINITION   AGENCOURT_8483154 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6305997
             5', mRNA sequence.
ACCESSION    BQ712831
VERSION      BQ712831.1  GI:21851730
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 1269)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Susan L. Sullivan, PhD
             cDNA Library Preparation: ResGen, Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
```

```
                    Clone distribution: MGC clone distribution information can be
                    found through the I.M.A.G.E. Consortium/LLNL at:
                    http://image.llnl.gov
                    Plate: LLAM13721  row: c  column: 22
                    High quality sequence stop: 264.
FEATURES            Location/Qualifiers
source              1..1269
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /clone="IMAGE:6305997"
                    /clone_lib="NIH_MGC_129"
                    /lab_host="DH10B (phage-resistant)"
                    /note="Organ: olfactory epithelium; Vector:
                    pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2: NotI; Cloned
                    unidirectionally. Primer: Oligo dT. Average insert size
                    2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this
                    is a NIH_MGC Library."
BASE COUNT    120 a    560 c    334 g    254 t     1 others
ORIGIN

US-09-677-653A-50 (1-647) x BQ712831 (1-1269)

Alignment Scores:
Pred. No.:                1.15
Score:                  111.50
Percent Similarity:     38.27%        Length:          1269
Best Local Similarity:  26.71%        Matches:           74
Query Match:             3.30%        Conservative:      32
DB:                        14         Mismatches:       101
                                      Indels:            70
                                      Gaps:              15

Qy   223 ProThrThrTyrArgAspValProAspProThrGluGlyLeuValArgThrValSerAspTyrArg 242
         ||| |||                                      ||| ||
Db    15 CCTGCACAGCCATGTCCATCCTTCTACGTCTCTTCCTCACCCGACGCCTTCC-------------- 65

Qy   243 LeuThrThrTyrLysAlaIleThrCysGlnAlaAsnMetPro---ThrLeuValAspGlnGly 261
         |||                                     ||| ::||| |||
Db    66 ----------------CCAGCCTCCGAGCCCTCATCGCTGCCCGATACGGGG 101

Qy   262 PheThrGlnIle----------GlyGlyGlnTyrAlaLeuThrProThr------SerLeu 275
         :::                   |||||||| ||  ||          ||
Db   102 AAGCTGGTGACGGTCCCGGGTGGGGAG---GCCCTCACCCCCCATCTGCTTGCAGCCTC 158

Qy   276 ProGlnTyrAspValSerGluAlaTyrAlaLeuHisThrLeuThrPheAlaLArgProSer 295
         |||:::      |||               ||||::::::|||       ||| |||
Db   159 CCCCGA----------GCAGCCGGACTCCATTCCCTCCGCCTCGCCCTTGCCCGCC 203

Qy   296 SerAlaAlaAlaLeuAlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProAla 315
         |||||||                   |||   |||          ||  ::
Db   204 TGGAGCAGGAGACCTGGCGCCCGTGTGGGGTGTGGGGCCCCGGCTGTACCTCAGCTGTTGT 263

Qy   316 GlyThrThrProAlaTrpGluGlnAlaSerSerGlyGlyTyrLeuThrTrpArgHisAsnGly 335
         :::              |||||   ||||               |||  ::  |||
Db   264 GGCCTGCAGGCCTGGGGAGGACCCGGGGCACCCTGTCCCTGGTCCATCAGCGGG 323

Qy   336 ThrThrPhePheProAlaGlyGlySerValSerTyrValLeuPro-GluGlyPheAlaLeuGluAr 355
         |||  |||||   ||||||                |||
Db   324 TCAGTCATGCTGACACGGAGC---TCATACCCGCTCGCCTGCGGGGC 366

Qy   355 gTyrTyrAspPro----------AsnAspGlySerTrpThrAspPhe------------ 366
         ||||:|:                 ||||||
Db   367 ----GACCTGCCTGCCCTGGGAATTTAAGGTCCTGGGTAGGAGCCCCAGGGTGGCCTTG 422

Qy   367 ----AlaSerAlaGlyAspThrValThrPheArgGlnValAlaValAlaAspGluValValVa 385
         |||||||                       ||::||::::
Db   423 GGGCCCTGGGGCGCAGGCCTTGAAACCCCCTGAGG----------CCCGGCTGCGCCCCT 470

Qy   385 lThrAsnAsnProAlaGlyGlyGlySerAlaProThrPheThrValArgValProProSe 405
         |||                 |||||    ||    ||| |||
Db   471 CGACACCTACCCCGGCGGGGGGATGCCCCCCCCCCCTCGTCTTCCCCTCGCCTCCTCCT 530

Qy   405 rAsnAlaTyrThr----------AsnThrValPheArgAsnThrLeuLeuGluThrAr 421
         ||| |:::                ||::||| |::::|||
Db   531 AGCCCGTTATTCCCCGACTACTCTACCCCGCAGCTCGCTTTCG------------ 582
```

```
Qy   421 gProSerSerArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrValAl 441
         ||||:||::|::|||||
Db   583 --------------------TTGCCAATCCCCTCCG------------ 600

Qy   441 aAsnAsnProLysIleGluGlnSerLeuLeuLysGluThrLeuGlyLeuCysTyrLeuValHi 461
         |||                       :::||| ||
Db   601 ---TTTCCGTGGGCCCGGCGCCCGGCGCGATTTTCCATCCTCCGGGGTCGCCTAGTCC 656

Qy   461 sSerLysSerMetArgAsnProValPhe------GlnLeuThrProAlaSer 475
         |||||:|||:||||                   ||::|||
Db   657 CACCCGGATTCGGTCTCCCCCCGCGTGTCCCCGAGCAGGTGTCCCCGGAAT 705

RESULT 6
BI082474
LOCUS       BI082474         2128 bp    mRNA    linear     EST 20-JUN-2001
DEFINITION  602877815F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5009474 5',
            mRNA sequence.
ACCESSION   BI082474
VERSION     BI082474.1   GI:14500804
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2128)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11055  row: n  column: 03
            High quality sequence stop: 42.
FEATURES            Location/Qualifiers
source              1..2128
                    /organism="Mus musculus"
                    /strain="FVB/N-3"
                    /db_xref="taxon:10090"
                    /clone="IMAGE:5009474"
                    /clone_lib="NCI_CGAP_Mam2"
                    /dev_stage="5 months"
                    /tissue_type="tumor, biopsy sample"
                    /lab_host="DH10B"
                    /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                    Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                    Library constructed by Life Technologies. Investigator
                    providing samples: Gilbert Smith, NIH"
BASE COUNT   576 a    872 c    330 g    350 t
ORIGIN

US-09-677-653A-50 (1-647) x BI082474 (1-2128)

Alignment Scores:
Pred. No.:                4.58
Score:                  109.50
Percent Similarity:     35.89%        Length:          2128
Best Local Similarity:  22.70%        Matches:           74
Query Match:             3.25%        Conservative:      43
DB:                        13         Mismatches:       133
                                      Indels:            76
                                      Gaps:              16

Qy   250 CysGluAlaAsnMetProThrLeuValAspGlnGlyPheTrpIleGlyGlyGlnTyrAla 269
         |||  ::::::|||||::
Db   352 TGCCGGCCGCTTCAATTCCCACCATTCCCTGTCCTGGACTCACCACGGGAAACCAATGGAAC 411

Qy   270 Leu-----ThrProThrSerLeuProGlnThrTyrAsp------------ValSerGluAlaTyr 284
         :::                                   |||::::||| |||
Db   270 Leu
```

```
RESULT 7

BC010575
BC010575
LOCUS       BC010575      3572 bp    mRNA    linear   HTC  12-JUL-2001
DEFINITION  Homo sapiens, Similar to E74-like factor 1, clone IMAGE:4338775,
            mRNA.
ACCESSION   BC010575
VERSION     BC010575.1  GI:14714847
KEYWORDS    HTC.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3572)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (10-JUL-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lou Staudt
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalon@bcm.tmc.edu
            Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B, Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.
            This clone has the following problem: Incomplete processing.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 12 Row: O Column: 13
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Genomescan gene prediction, Similarity but not identity
            to protein
FEATURES             Location/Qualifiers
     source          1..3572
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4338775"
                     /tissue_type="lymph, lymphoma"
                     /clone_lib="NIH_MGC_85"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT   1135 a   678 c   703 g   1056 t
ORIGIN

US-09-677-653a-50 (1-647) x BC010575 (1-3572)

Alignment Scores:
Pred. No.:
Score:                     12.6
                         109.00
Percent Similarity:       35.22%          Length:        3572
Best Local Similarity:    20.71%          Matches:        117
Query Match:               3.23%          Conservative:    82
DB:                          11           Mismatches:     214
                                          Indels:         153
                                          Gaps:            22
```

```
Db  412 ATGATGGAACGTGAAAACAATTACCAAACTACACGCCCGGCGTCGGTCACGGAGGATTAC 471
Qy  285 -------------------------------AlaLeuHisThrLeu-ThrPheAlaArgPr 294
Db  472 CACACGGAAAAGGGGCGGGACAACAACGGGGGGCGGAATCACCTCCTTCATCACCCCGCC 531
Qy  294 oSerSerAlaAlaAlaLeuAlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaPr 314
Db  532 AGTTACTACTCCCCCATC----------------ATTCCACCGCCCTGGTTCTAAGCC 573
Qy  314 oAlaGlyThrProAlaTrpGluGlnAlaSerSerGlyGlyGlyTyrLeuThrArgArgHis-- 333
Db  574 CCCATCCTCCCTTCCACCATATCAGCCCCCCGCTAAAGGCAGAACATAATCAATTCGGTCAAC 633
Qy  333 ---AsnGlyThrThrPheProAlaGlySerValSerTyrValLeuProGluGlyPheAl 352
Db  634 ACCCAACACGGCATGTCAACCCCACGGTCGCACAAATCCATTGGATCCCCCCTTGGGACACT 693
Qy  352 aLeuGluArgTyrTyrAspProAsnAspGlySerTrpThrAspPheAlaSerAlaGlyAspTh 372
Db  694 CCTTAACCACTTT--- ... AACCGCAGCGCGATAC 720
Qy  372 rValThrPheArgGlnVal------AlaValAlaSerGluValValValValThrAns 387
Db  721 ACTGGGCGGACGTCGGTACCACACTAGGAACGCACTTATACCGCCATCAACGCATATCA 780
Qy  387 nAsnProAlaGlyGlyGly------SerAlaProThrPheThrValArgValProPheAl 405
Db  781 TTACCCTACTTAGGCGGTAATTCCATCCCCATTAGTCACAAATTTCACTTCACTCACCCG 840
Qy  405 eGlyGlnThrValAlaAsnAsnProLysIleGluGlnSer-----LeuLeuLysGluTh 419
Db  841 CCCTACACATAGTCACCTTGGCGGCACTCACCTCCCACATACGTACGTCACTACGTCACCACGA 900
Qy  419 uThrArgPro-----SerSerArgArgLeuGluLeuProMetProProAlaAsppPh 436
Db  901 GGCAAAACCCCCAAACCCTTTCCACAACCATTACCACCCTGTCATACCCCCCCTT----- 955
Qy  436 eGlyGlnThrValAlaAsnProLysIleGluGlnSer------LeuLeuLysGluTh 454
Db 1012 CCTTACCATACAC----- ... CACACCTTCTAACCACAACCTTCTCACCCGTPACTCCCAC 1065
Qy  454 rLeuGlyCysTyrLeuValHisSerLysSerMetArgAsnProValPheGlnLeuThrProAl 474
Db 1066 ACATCCTTTCAACGCTTGGACAATTACTCCAAGTCCTTATCCCGCGTCGACACGGTACC 1125
Qy  474 aSerSerPheGlyAlaValSerPheAsnAsn--------------------Pr 485
Db 1186 ATGACTAAACCACA--- ... CGCTCTCCATCTCAATCCCCATC-- 1222
Qy  505 nAsnMetSerThrAlaAlaAlaHisPheArgSerLeuSerHisSerCysSerIleValTh 525
Db 1126 AGCCGTACCCGCAGACAACCGGATAATCCCCCAGGCGTGCACAGCGGACAAAACG 1185
Qy  525 rLysThrTyrGlnGly 530
Db 1223 ----ACCTACTACGGC 1234

Qy   97 LysTyrLeuAspProAlaGlyAlaThr--------GluSerAlaAlaValGlyGlu 113
Db  781 AAATATGCAGAGACTCACCGGGAGCCTCATCACCAGAAGAGGAAAAAGGAAGA 840
Qy  114 TyrSerLysIleProAspGlyLeuValLysPheSerValAspAlaGluIleArgGluIle 133
Db  841 AAAACTAAACCACA----------- 855
Qy  134 TyrAsnGluGluCysProValValThr---AspValSerValPro------LeuAsp 149
Db  856 ---CGACCAGATTCCCCAGCCACTACCCCAAAATATATCTGTGAAGAGAAAACAAAGAT 912
Qy  150 GlyIleArgGlnTrpSerLeuSerIlePheSerPheProMetPheArgThrArgTyrValAla 169
Db  913 GGAAAGGGAAACACAATTTAATCTTTGGGAGTTT-----------TTA 948
Qy  170 ValAlaAsnValGluAsnLysGluMetSerLeuAspValAlaAsnAspLeuIleGluTrp 189
Db  949 CTGGCACTGCTCCAGGACAAGGCT-------------ACTTGTCCTAAATACATCAAGTGG 996
```

```
Qy   190 LeuAsnAsnLeuLeuAlaAspTrpArgTyrValValAspSerGluGln--------Trp 205
              |||:||||||
Db   997 ACCCAGGACGAGAAAAGGCAAATTTTAAATGGTGGAATCTCAAAGCAGTGTCCAGGTTGTGG 1056

Qy   206 IleAsnPheThrAsn------AspThrThrTyrTyrValArgIleArgValLeuArgPro 223
             |||       :||  ||| |||||||
Db  1057 GGGAAGCACAAAAACCTGATATGAATTAATTAGAGACAGTGGAGGAGGACTCAGGTAC 1116

Qy   224 ThrTyrAspValProAspProThrGluGlyLeuValArgThrValSerAspTyrArgLeu 243
             |||
Db  1117 TATTACCAA---------AGGGGTATTCTCGGCAAAAGTGGAAGGTCAGGCCTTG 1161

Qy   244 ThrTyrLys--------- Ala 247
             |||:||                      :::
Db  1162 GTGTATCAGTTTAAAGAAAATGCCAAAAGACTTATAATATATATAAAAATGATAGGAGGATCCAAGT 1221

Qy   248 IleThrCysGluLeuAlaAsnMetProThrLeuValAspGln--------- 260
             |||:|||  ||| 
Db  1222 TCCAGCATAGAGTCTTCAGATCCATCGCTATCTTCATCACCACTTCAAATAGGAATCAA 1281

Qy   261 ---------GlyPheTrpIleGlyGlyGlnTyrAlaLeu 270
                             |||          :::
Db  1282 ACCAGCCCGGTCGGAGAGTATTCTTCAAGTCCAGGGGTAAAAGGAGGAGCCACTACAGTTCTA 1341

Qy   271 ThrProThrSer-LeuProGlnTyrTyrAspValSerGluAlaTyrAlaLeuHis-ThrLeuT 290
             |||       :::     |||          |||::  |||
Db  1342 AAACCAGGGAATTCTAAAGCTGCAAAACCCAAAGATCCGTGTGGAAGTTGCACAACCATCA 1401

Qy   290 hrProPheAlaArgProSerSerAlaAlaAlaAlaAlaLeuAlaPheValTrpAlaGlyLeuProGlnG 310
             |||:||          |||||||
Db  1402 GAGGGAGGAAAGCAGCCTAGAGAACCAGTACCAATGCCAGGATGAAAACATTAAAATCTTCCGTTCAG 1461

Qy   310 luGlyPheAlaLeuGluArgTyrArgAspProAsnAspGlySerTrpThrAspPheAlaSerA 330
             |||                         :::           |||
Db  1462 GTTCATGTAGTAGTACAGCCAGTACA---GGCTGTCCCA 1494

Qy   330 hrArgProArgHisAsnGlyThrThrPheProAlaGly---SerValSerTyrValValLeuProG 349
             ||||||||||                   |
Db  1495 GAGGGAGGAAAGCAGCCTAGAGAACCAGTGAAACAATTAAAATCTTCCGTTCAG 1554

Qy   349 laGly-AspThrValThrPheAlaGlnValAlaValAlaAspGluValAlaValThrThrAsnAsn 369
             |||:||                                   |||
Db  1555 AGTATTAGGACTATACAACAGGCCTCCAACCAAGTTCCAGTGGTTGTTGTCTCCTAGGAATCAG 1614

Qy   369 ProAlaGlyGlyGlySerAlaProThrPheThrPheValArgValProProSerAsnAlaTyr 389
             |||||||      :::        |||       |||
Db  1615 CAGTTGCATACAGTAACACTCCAAACAGTGCCACTCACAAACAGTTATACCAGCACAGAT 1674

Qy   389 ThrAsnThrValPheArgAsnThrLeuLeuGluThrArgProSerSerArgArgLeuGlu 409
             |||            |||              |||
Db  1675 CCATCACAGGGTACTCAGAGAGTTTATTTTACNAGCCATTCCATCACACAGCC 1734

Qy   409 ---------
             ||||||||||
Db  1735 ATGACAGTACTGAAAGAAANTGTCAGTCACAAAAGGGGCTCCCTCCTTCA 1794

Qy   429 LeuProMetProProAlaAspPheGlyGlnThrValAlaValSerPhe--------- 448
             |||      :::                          |||:||
Db  1795 ATTGTCTTGGGCCCTGCCAGGTTCAGCAGGTCCTTACTAGCAAT------GTTCAG 1845

Qy   449 SerLeuLeuLysGluThrLeuGlyCysTyrLeuValHisSerArgAsnProVal 468
             |||                    :::
Db  1846 ACCATTTGCAAAGGAACCGTC 1887

Qy   469 PheGlnLeuThr--ProAlaSerSerPheGlyAlaValSerPhe 482
             |||       :::                       :::|||
Db  1888 TTCAGTGCTACTGCACCTGGTGGTCACAGTGTGGTTGCTCA 1946

Qy   483 ------------AsnAsnProGlyTyrGluArgThrArgAspLeuProAspTyrThr 497
                          ::::|||::
Db  1947 CCCACCTTGGCACTGTTAATCACTTCAGTTATCAAAGAAAACAAAAACTCTTACA 2006

Qy   498 GlyIleArgArgAsp-------SerPheAsp 508
             |||::                       :::
Db  2007 GGAAGTAGAAGAAAAAAAGGAATCTGAAGAGATCACATTTGAAGAGAAGACACTGAGAAAAGCGGAGCA 2066

Qy   509 ThrAlaValAla-------HisPheArgSerLeuSerHisCysSer 522
             |||  |||                    :::||
Db  2067 GCAGCCACAGCCCTTAGTGTGAGGTAGTGTCCAGTTCCAATGGATTGTCTCTCAGTTACTGC 2126

Qy   523 IleValThrLysThrThrTyrGlnGlyTrpGlnGlyValThrArgAsnValAsnThrPro----- 540
             ||||||||||            :::                  |||||||
Db  2127 TATGAAAACAAAAACGAACTGCCTGGAACCCAA---CTCTTTTTAGTTAATAATACCAAAGCTT 2183

Qy   540 ---------
                      :::
Db  2184 ATGAAATAAATTGTTTGTTTGTTTAATTGAACATTTTCAATTATATAGCAGACTGACTGATTCTAAGA 2243

Qy   540 ---------
                       :::
Db  2244 TAAATCTAAGGAGGTTTCTAAATTTGTAATTGTTAAAAATGAGAGTTAATTTGACTTTG 2303

Qy   541 -----PheGlyGlnPheAlaHisAlaGlyLeuLeuLeuArgSerAsnGluLeuIleLeuCysLeu 558
             |||||||||||||||||||:::|||

Qy   559 AlaAspAsp 561
             :::|||:::
Db  2304 TTAGATGAG 2312
```

RESULT 8

BI413164

LOCUS   BI413164   915 bp   mRNA   linear   EST 14-AUG-2001

DEFINITION 602990033F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5145873 5',
           mRNA sequence.

ACCESSION   BI413164

VERSION   BI413164.1 GI:15174087

KEYWORDS   EST.

SOURCE   house mouse.

  ORGANISM Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1 (bases 1 to 915)

  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.

  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)

  JOURNAL   Unpublished (1999)

COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Gilbert Smith, Ph.D.
           cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
           Bonaldo, Ph.D.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM11360 row: a column: 10
           High quality sequence start: 5
           High quality sequence stop: 821.

FEATURES   Location/Qualifiers
     source   1..915
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5145873"
                     /clone_lib="NCI_CGAP_Lu33"
                     /tissue_type="pooled lung tumors"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
                     strand cDNA was prepared from mRNA obtained from pooled
                     lung tumors with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGAGCGGCGCGCGCCGCTCTCGTTTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT   218 a   268 c   267 g   162 t

ORIGIN

US-09-677-653a-50 (1-647) x BI413164 (1-915)

Alignment Scores:

Pred. No.:	1.37	Length:	915
Score:	108.50	Matches:	80
Percent Similarity:	36.578%	Conservative:	33
Best Local Similarity:	25.89%	Mismatches:	115
Query Match:	3.228	Indels:	81
DB:	13	Gaps:	14

```
Qy 228 ProAspProThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyr------ 245
       ||||||          |||||||:|||||||::||||||:::
Db 119 CCTGACAGCCTGGAGGTACTGGTGAAGACCCTGGACTCCGACTACCGGGACTTTATTGTG 178

Qy 246 ------------------------LysAlaIleThrCysGluAlaAsnMetPro 255
                                ::: |||
Db 179 GGGGCCCAGATGAATGAGTAAAGGAGTTTAAGGAACACATAGCTGCCTCTCGTCAGCATCCCT 238

Qy 256 ThrLeuValAspGlnGlyPheTrpIleGlyGlnTyrAlaLeuThrProThrSerLeu 275
       ::                  :::|||
Db 239 TCCGAGAAACACGGGCTCATCTACCAGCCGGTCCTACACAGACAGCAAGAAGCTCC--- 295

Qy 276 ProGlnTyrAspValSerGluAlaThrAlaLeuHisThrLeuThrPheAlaArgProSer 295
       ::||:|||     :::|||     :::|||
Db 296 -----AGGAACGTTGGG--GGAAAGGTTATTCACCTGGTGGAACGGGCTCCTCCTCAG 346

Qy 296 SerAlaAlaAlaLeuAlaPheValTrpLeuProGlnGlyGlyThrAlaProAla 315
       :::                       ||||||
Db 347 ACTCAG----------------------CTTCCTTCTGAGCATCTTCTGGGACA 379

Qy 316 GlyThrThrProAlaTrpGlnIleAlaSerSerGlyGlyTyrLeuThrThrArgHisAsnGly 335
       |||:::                ::               ||||||
Db 380 GGCTCTGCCTCA------------GCAACTACATGGTGGGGGGCACCCCTG------CCTGGC 421

Qy 336 ThrThrPheProAlaGlySerVal-------------------SerTyrValLeuProGlu 349
       ||||||                                   |||||||:::
Db 422 ACTCGGGGCGCCCTGGGGCTTCTGTTCATGACGGGAATGCCAACAGCTATGGTACTGGTGGA 481

Qy 350 GlyPheAlaLeuGluArgTyrAspProAsnAspGlySerTrpThrAsp-PheAlaSerAl 369
       ||:::|||||||||||                  :||  |||  |||
Db 482 ACCTTCAATCTT------------CCTAGTGACCGGCTCTGCTGTGGATGTTCACATCAAC 529

Qy 369 aGlyAspThrValThrPheArgGlnValAlaValAlaAspGluValValThr-AsnAsnP 389
       :::                            |||||
Db 530 ATGGAACAGGCCCCAATTCAGA------------------GTGAGCCCCGGGTC 565

Qy 389 roAlaGlyGlyGlyGlySerAla-------------------ProThrPheT 399
       |||:::                                  |||
Db 566 CGGCTGGGTGATGGCTCAGCACATGAATATACAGACCTTACTGTGCCCGGATGGAG 625

Qy 399 hrValArgValProProSerAsnAlaThrValPheArgAsnThrLeuLeuG 419
       |||                                      ||||||:::
Db 626 TGTCGAGGGGAGGGGAACCAGACTCAGCCACCCCCGCAGACAC-------- 675

Qy 419 luThrArgProSerSerArgLeuGluLeuProMetProProAlaAspPheGlyGlnT 439
                                                ||||||||
Db 676 --------------------------CGCAGACTGTGGCCCTCAGAGACGGGAAGCCTTGAACTCACAA 718

Qy 439 hrValAlaLaAsnAsnProLysSerIleGlnSerLeuLeuLysGluThrLeuGlyCysTyrL 459
       ||||||||            ::
Db 719 ACATCAGAAAACCAGTCGAAAGTG---AAGCTCCTCCTCGAAGAGACCCATGGGA------- 768

Qy 459 euValHisSerLysMet------------------------ArgAsnProValPheGlnLeuT 472
       :::||                                   |||
Db 769 --GTCAAGGAAGAAAATGGAGGGGAACGGGTCCCCAAACCAGAACTCCAGAGACTTCGCGCCCT 826

Qy 472 hrProAlaSerSerPheGlyAlaValSerPheAsnAsnProGlyThrGluArgThrArgA 492
       |||||||                                ::
Db 827 CAAGGACCAGCTCCCGGCAGGCCCCAGCTCCCGGCGCCCAGCGCTCCCGGCGCGAGAGAC 886

Qy 492 spLeuProAspTyrThrGlyIle 499
       ||||||||||
Db 887 AAAGTGCGCGCCCAGAACAGGCATT 909
```

US-09-677-653A-50 (1-647) x BM926649 (1-1776)

Alignment Scores:

Pred. No.:	4.29	Length:	1776
Score:	108.50	Matches:	77
Percent Similarity:	39.06%	Conservative:	23
Best Local Similarity:	30.08%	Mismatches:	106
Query Match:	3.228	Indels:	50
DB:	14	Gaps:	12

```
Qy 306 GlyLeuProGlnGlyGlyThr-AlaProAlaGlyThrThrProAlaTrpGluGlnIleAlaSerSe 325
       |||                   ||||||||||||||||||
Db 65  GGCTATCCTCCTGGTGGAACGGGCTCCTCCTCCACACTCACCTCCCTTCTGGGCGATCTTC 124

Qy 325 rGly---GlyTyrLeuThrTrpArgHisAsn------------GlyThrThrPhePr 339
       |||     |||                              ||||||
Db 125 TGGGACGGGGTCGTGCCTCACCACTCATGTGTGGGGGGGATCCCCCTGGTACTCGGGGC 184

Qy 339 oAlaGlySerVal--------------------SerTyrValLeuProGluGlyPheAlaLe 353
       :::                             ||||||||||:::
Db 185 TGGGGCCTTCTGTTCATGACCGGAATGCCAACAGCTATGGTTGGTGGAACCTTCAAATCT 244

Qy 353 uGluArgTyrAspProAsnAspGlySerTrpThrAsp-PheAlaSerAlaGlyAspThrV 373
       ::::                                  |||  :::|||
Db 244 TGGGGCCTTCTGTTCATGACCGGAATGCCAACAGCTATGGTTGGTGGAACCTTCAAATCT 373
```

```
FEATURES             Location/Qualifiers
     source          1..861
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4415156"
                     /clone_lib="NIH_MGC_89"

LOCUS       BG168931      861 bp    mRNA    linear   EST 06-FEB-2001
DEFINITION  602202811F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4415156 5',
            mRNA sequence.
ACCESSION   BG168931
VERSION     BG168931.1  GI:12675634
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 861)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10144 row: b column: 21
            High quality sequence start: 2
            High quality sequence stop: 772.
            /tissue_type="hypernephroma, cell line"
            /lab_host="DH10B (phage-resistant)."
            /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.3 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
BASE COUNT      167 a    284 c    246 g    164 t
ORIGIN
```

```
RESULT 10
BG168931
BG168931

Alignment Scores:
Pred. No.:              1.4           Length:        861
Score:                108.00         Matches:        71
Percent Similarity:    36.40%        Conservative:   28
Best Local Similarity: 26.10%        Mismatches:     93
Query Match:            3.20%        Indels:         80
DB:                    12            Gaps:           16
```

```
us-09-677-653a-50 (1-647) x BG168931 (1-861)
```

```
Qy    245  T---------------------CCTAGTGACGGCTCTGCTGTGGATGTTCACATCAACATGGAACAGGC  292
                                  |||:::||||||||||||||
Db    293  CCCGATTC------------------------ATAGTGAGCCCCGGGTACGGCTGGTGAT  328

Qy    293  CCCGATTC----------ATAGTGAGCCCCGGGTACGGCTGGTGAT  328

Qy    373  alThrPheArgGlnValAlaAlaValAspGluValValThr-AsnAsnProAlaGlyGly  392
                  ||:       :||     ||||||:       |||||||
Db    365  ATCCCGGATGGAGTGTCGAGGAGGGCCCCAACCGACACCAGTCAGCCGCCCCGCACC  424

Qy    393  GlySerAlaProThrPheThrValArgValProProSerAsnAlaTyrThrAsnThrVal  412
                  |||||||||         ||||||::    :|
Db    329  GGCTCACCA----------CATGATCAGGGATATACAGACTTACT  364

Qy    413  PheArgAsnThrLeuLeuGluThrArgProSerSerArgArgLeuGlyLeu-ProMetPr  432
                  |||          ||:||       ||||||:
Db    485  TGAAGCAACCTCCCCGGGAGCCCATGGAGGCGCAGAAGAAGTGGAGGAAC---------G  532

Qy    432  oProAlaAAspPheGlnThrValAlaAlaAsnAsnPro-LysIleGluGlnSerLeu--L  451
                  |||||||    |||    ::|       |||:       :
Db    425  ----------CATGATCAGGGATATACAGACTTACT  364

Qy    451  euLysGluThrLeuGlyCysTyrLeuValHisSerLysSerMetArgAsnProValPheGlnLeu  471
                  ::|||     |||       |||||:||||||||
Db    471

Qy    471  eurThrProAlaSerSerPheGlyAlaValSerPheAsnAsnProGlyTyrTyrGluArgThrA  491
                  ::|       |||:||||:||       ||||||||
Db    533  TGCCCCCACCCCAGAACCCGAGCTCACTCCTGCCCCACCGGGCCCAACACC-  591

Qy    491  rgAspLeuProAspTyrThrGlyIleArgAspSerPheAsp---GlnAsnMetSerThrA  510
                  |||       |:     |||||||||
Db    592  ------TGCCCCGGAAACAAAATGCCACCACCAATCCTTCCCTGGGGAATATGTCGAGGT  646

Qy    510  laValAlaHisPheArgSerLeu-------SerHis----------SerCysSerIleValT  525
                  :||       ||||||       |||       :::
Db    647  GCTCCAGGAGGTACAGGGGGTGGAAAGTCACCCTCCACCCCTTCTTGCAGCTCTACCA  706

Qy    525  hrLysThrTyrGlnGlyThrTrpGluGlyValThrAsnVal  537
                  |||       |||       |||
Db    707  ACAAGGGTCTCGGGGGGTCGGGGCTGGCCGTCGACCACCAGGACTA  744
```

```
Qy    263  TrpIleGlyGlyGln--------------------TyrAlaLeuThrProThrSerLeu------ProGln  277
                  |||                 ||:||       ::|||
Db    108  TGGTTTGCTGGGGACGAACGCGCTCCCGAGCCGCGTGTTTCTTCCACGTCGGCGGCGCCCAGA  167

Qy    278  TyrAspValSer---GluAlaTyrAlaLeuHisThrLeuThrPheAlaArgProSerSer  296
                  :::       |||||||::|||       |||       |||
Db    168  CACCAGGCGGTCATGGGTGGCGGCATGGGCCCAGAAGGAGGACTCCTACG------  209

Qy    297  AlaAlaAlaLeuAlaPheValThrProAlaGlyLeuProGlnGlyGlyThrAla------Pro  314
                  |||||||       |||:::       |||       |||
Db    210  ----------TGGGCGACGGAGGGCCCAGAACGAACGGTGGCATCCTGACC  248

Qy    315  AlaGlyThrProAlaTrpGluGlnAlaSerSerGlyGlyThrTyrLeuThrPhrArgHisAsn  334
                  |||||||       |||||||       ||:::
Db    249  TGAAGTACCCATTG---AGCATGGCATGGTCACCAACTGGGACGACAGGGACGAAGAGATCT  305

Qy    335  GlyThrThr--------------------PheProAlaGlySerValSerTyrVal---  346
                  |||||||||                 |||::
Db    306  GGCACCACACCTTCTACAACGAGCGTGCGCGTGGCCCCGGAGGAGCCACACCCAGTGCTGCTGAC  365

Qy    347  --------LeuProGluGlyPheAlaLeuGluArgTyrArgAspProAsnAsp----------  360
                  ||||||||       |||       :::|||
Db    366  CGAGGGCCCCCTGAACCCCAAGGCCAACACAGAGAGAAGATGACTCAGATTATGTTGAGAC  425

Qy    361  ------------------------------------------------------------Gly  361
                  :::
Db    361

Qy    362  SerTrpThrAspPheAla----SerAlaGlyAspThrValThr----------------Phe  375
                  |||       :::|||||||       |||
Db    486  CGCACCACTTACCTCATGAAAGATCCCTCACTGAGCCGAGGCGTCACCACACGGTGCCCATCTAC  545

Qy    376  ArgGlnValAlaValAspGluValValThrAsnAsnProAlaGlyGlyGlySerAla  395
                  |||||||       |||||||:::       |||:::
Db    546  GAGGGCTACGGCCTCCCCACCACGCCATCTCTGCGTCTTGGACCTGGCCGCCGGGACCTGA  605

Qy    396  ProThrPheThrValArgValProPro---SerAsnAlaTyrThrAsnThrValPheArg  414
                  |||||||::::       :::|||       |||
Db    606  CCGACTTACCTCATGAAAGATCCCTCACTGAGCCGAGGCGTCACCACACGGTGCCCATCTAC  665

Qy    415  AsnThrLeuLeuGluThrArgProSerSerArgArgLeuGlyLeuProMetProProAla  434
                  |||||||||||       |||       |||||||
Db    666  GAGGGCTACGGCCTCCCCACCACGCCATCCTCGGTGCCGCCGACCATCCAAGGAGAAGCTTTGGCTA---  722

Qy    435  AspPheGlyGln-ThrValAlaAsn---AsnProLysIleGluGlnSerLeuLeuLysGl  453
                  ||||||       |:::       |||       :::
Db    723  GACTTCGAGCAAGGAGAATTGGCAACCACCATCCTCTCTGGAGAAAAGCTAAGGA  782

Qy    453  u------------------ThrLeuGlyCysTyrLeuValHisSerLysMetArgAs  466
                  |||||||
Db    783  GTTGCCCGATGGCCCAGGTTGTTCACCCTTGGC----------------ATGAACGG  821

Qy    466  nProValPheGlnLeuThr-ProAlaSerSer  476
```

```
Db   822  TCCGGTGTCCGAGCCGGTGTCCCGCTTCTTCC  853
          |||||||||||||||||||
Qy   ...  ...CCGGTGTCCGAGCCGGTGTCCCGCTTCTTCC  853
```

RESULT 11

LOCUS   AL578091/c

DEFINITION AL578091 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK003YI23 3
            prime, mRNA sequence.

LOCUS   AL578091   869 bp   mRNA   linear   EST 16-FEB-2001

ACCESSION   AL578091

VERSION   AL578091.1 GI:12941832

KEYWORDS   EST.

SOURCE   human.

ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 869)

AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE   Full-length cDNA libraries and normalization

JOURNAL   Unpublished (2001)

COMMENT   Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES   Location/Qualifiers
     source   1..869
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DK003YI23"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"

BASE COUNT   143 a   268 c   280 g   174 t   4 others

ORIGIN

Alignment Scores:

Pred. No.:   1.6

Score: 107.50

Percent Similarity: 35.29%   Length: 869

Best Local Similarity: 24.79%   Matches: 59
                                 Conservative: 25
Query Match: 3.19%   Mismatches: 90
                                 Indels: 66
                                 Gaps: 8

DB:

US-09-677-653A-50 (1-647) x AL578091 (1-869)

```
Qy   260  GlnGlyPheTrpIleGlyGlyGlnTyrAlaLeuThrProThrSerLeuProGlnTyrArgAsp  279
          |||  |||||  ||||||||||||  ||||||||||||  |||
Db   584  CAGAACTTCTGGATC------------------------AGCCTCCAGGAGACACTG  552

Qy   280  ValSerGluAlaTyrAlaLeuHisThrLeuThrPheAlaArgProSerSerAlaAlaAla  299
          |||||||  ||||                      |||
Db   551  TGCAGTGAGGAGATAGGCCCTGAGCACTGCCAGTGANGACCGC----------------  510

Qy   300  LeuAlaPheValTrpAlaGlyLeuProGlnGlyThrAlaPro---AlaGlyThrPro  318
          ||||  |||||||  ||||||||:::|||  :::
Db   509  ---------TGCTGGAACGGGATCCCAGAGGCCGGTACCTCCCCGAGGTGCATGGGTGACG  459

Qy   319  AlaTrpGluGlnIleAlaSerSerGlyTyrGlyTyrLeuThrTrpArgHisAsnGlyThrThrPhe  338
          |||||||      |||:::||                  |||
Db   458  GCCTGGCCAACCAGATCAACAACCCGAGGTGGACATCACCAAGCCGGACACATGA  399

Qy   339  ProAlaGlySerValSerThrValLeuProGluGlyPheAlaLeuGluArgTyrAspPro  358
          |||::|||:||||||||      |||
Db   398  CCATCCGGCCAGCTGTCCT----------------------GAACCGACTGACCCT  366
```

RESULT 12

LOCUS   AG139016

DEFINITION Pan troglodytes DNA, clone: PTB-153F24.R, genomic survey sequence.

LOCUS   AG139016   686 bp   DNA   linear   GSS 04-NOV-2001

ACCESSION   AG139016

VERSION   AG139016.1 GI:16668694

KEYWORDS   GSS.

SOURCE   Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-153F24.R.

ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE   1

AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE   BAC end sequences of Library PTB

JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 686)

AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE   Direct Submission

JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://ngp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT   Clones are derived from the chimpanzee BAC library PTB This BAC end
            clone tracking errors.
            was generated during the R&D process and may have higher chance of

PRIMERS   Sequencing: M13Rev

LIBRARY   Vector   : pKS145
            R.Site 1   : SacI
            R.Site 2   : SacI

FEATURES   Location/Qualifiers
     source   1..686
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-153F24.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"

BASE COUNT   158 a   269 c   117 g   142 t

ORIGIN

```
Qy   359  AsnAspGlySer----------------------------------------------TrpThrAspPhe  366
          |||||||||||                                               :::
Db   365  GAGGAGGCCGGCTTAGTGCCTGCTTGCTTTGCTTTTCATCACCGTCCCGCCACAGTGGACGGAGGTC  306

Qy   367  AlaSer------AlaGlyAspThrValThrPheArgGlnValAlaLeuAlaAspGluValVal  384
          :::||||     :::||||                              :::  |||:::|||:::
Db   305  CCCGGTGTCGCTGGTCAGGTCCCCATGGCTTGTTCTCTGGAACCTRACTTTA-GATGTTTTG  247

Qy   385  ValThrArgAsnAlaLeuProAlaGlyGlyGlySerAlaLeuProThrPheThrValArgValProPro  404
          :::             |||             |||       |||||:::||||||
Db   246  GGATCAGGAGCCCCAACACAGGCAAGTCCACCCCATA-ATAACCCTGCCACAGTGCCAGGG  188

Qy   405  SerAsnAlaTyrThrAsnThrValPheArgAsnThrLeuLeuGluThrArgProSerSer  424
          :::      |||      ||||                                     :::
Db   187  TGGGCTGCGGGGACTCTGCCACAGTGANG----------------CCGGCGCC  152

Qy   425  ArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrValAlaAsnAsnPro  444
          |||              ||||:::                          :::|||
Db   151  AGGAGCAGCAGAACTCCCGCTCCACACAGACGGGCCCTRAGGGGTGGGGCGCTCAGACCCACCT  92

Qy   445  LysIleGluGlnSerLeuLeuLysGluThrLeuGlyCysTyrLeuValHisSerLysMet  464
          |||                                |||
Db   91   ACGGTC----------------------ATCTCTGGAAGG  74

Qy   465  ArgAsnProValPheGlnLeuThrProAlaSerSerPheGlyAlaValSerPhe  482
          |||                             ||||||||||||||:::
Db   73   GGCAGCCCTGAGTGGTCACTGGTGCAGGGCAGTGGCCAAGCCTGCTGTGTCCTTC  20
```

Left column

```
Alignment Scores:
Pred. No.:            1.37
Score:             106.50
Percent Similarity:  32.84%
Best Local Similarity: 23.62%
Query Match:         3.16%
DB:                  17

                    Length:     686
                    Matches:     64
                    Conservative: 25
                    Mismatches:  77
                    Indels:     105
                    Gaps:        14
```

US-09-677-653A-50 (1-647) x AG139016 (1-686)

```
Qy   242 ArgLeuThrTyrLys------AlaIleThrCysGluAlaAlaAsnMetProThrLeuValAsp 259
         ||||||||||        :::       |||
Db    12 CGACTCACTATCAGGGTAGCCCGTAGCATGCTGGAGAGCCTCCCACTTTCCTTGTGACA 71

Qy   260 GlnGlyPheTrp-----------IleGlyGlnTyrAlaLeuThr 271
         ||||||              ::           |||
Db    72 GGAAGGTTCTGGACCCAGGCGTCTTCGTCCCCAGGAGCACGTCGTCACCAAGCCTGGACAACA 131

y    272 ProThrSerLeuProGlnThrTyrAspValSerGluAlaTyrAlaLeuHisThrLeuThrPhe 291
         ||||||              |||  |||||
Db   132 GCCCAACGGCCTGCCCTCCTTTTTTTGTTCTTCCACAGCATAT----- 170

Qy   292 AlaAlaArgProSerSerAlaAlaAlaLeuAlaPheValTrpAlaGlyLeuProGlnGlyGly 311
         |||              :::
Db   171 -----TGGGCCTGGCCCGGGACAACCACACGGGCACA 200

Qy   312 Thr--AlaProLeuAlaGlyThrProAlaTrpGluGlnAlaLeuSerSerGlyGlyTyrLeuThrThr 331
         |||  |||||||                  |||
Db   201 ACTTGGCTCCTGGCT-----------CCGGATTGG 227

Qy   331 rpArgHisAsnGlyThrThrPheProAlaGlyLeuSerValSerTyrValLeuProGluGlyP 351
         ||:::                  :::|:::|:::
Db   228 GGAAGAGAAAT-----CCCCTGGCCTTCCTGGCCTTCCTT----- 266

Qy   351 heAlaLeuGluArgTyrArgAspAlaGlySerTrpPheAlaSerAlaGlyA 371
         |||                  |||
Db   267 TCCACCCCCCAAAAATTCAAACCGGTGAACGCTTTTGGAAACCCGGAAATTCTGGC---- 322

Qy   371 spThrValThrPheArgGlnValValValThrAsnAsnProAlaG 391
         :::                  |||||
Db   323 ----------CCAAAAAATCCTCCCT 338

Qy   391 LyGlyGlySerAlaProThrPheThrValArgValProProSerAsnAlaTyrThrAsnT 411
         |||||||||||                  :::|:::|:::
Db   339 CGGGGGGCTCCCCCAAAAAACACACGGGACACCACCGGATCTCAAGCGCCCTCCGCAAACC 398

Qy   411 hrVal------------PheArgAsnThrLeu- 417
         :::                  |||
Db   399 CCAtGTGTTATTATGGGCCAAAATGTGCGGTTCACACATTCCACCAGACCTCT 458

Qy   436 heGlyGlnThrValAlaAlaAsnAsnProLysIleGlnSerLeuLeuLysGluThrLeuG 456
         |||              |||  |||
Db   459 TTCCAAttAACACGCGCCCGAAATATCCAAAAGATTCCCTCGCCT 502

Qy   456 LyCysTyrTyrLeuValHisSerLysMetArgAsnProValPheGlnLeuThrProAlaSerS 476
         |||||||||                  |||||||||||
Db   503 -----CCCAAA-----CAACACCTC-----ACCTCCCGCG 527

Qy   476 erPheGlyAlaAlaValSerPheAsnAsnPro 485
         ||              :::|:::|||
Db   528 GATGCTACACC-----CAACTTACTCCGCCCTTCT 557

Qy   558 CCCACCTATCAAATTTCTCTCTCAGCACCG 586
         ||              ::|:||
         CCCACCTATCAAATTTCTCTCTCAGCACCG 586
```

RESULT 13
```
LOCUS       BI650823        820 bp   mRNA   linear
DEFINITION  603277718F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5338301 5',
            mRNA sequence.
ACCESSION   BI650823
```

Right column

```
VERSION     BI650823.1  GI:15565059
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE   1  (bases 1 to 820)
  AUTHORS   Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
            NIH-MGC http://mgc.ncl.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE     Unpublished (1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11859 row: c  column: 06
            High quality sequence stop: 729.

FEATURES             Location/Qualifiers
     source          1..820
                     /organism="Mus musculus"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5338301"
                     /clone_lib="NCI_CGAP_Mam3"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally; Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
                     /lab_host="DH10B"

BASE COUNT      122 a    289 c    249 g    160 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:            1.86
Score:             105.50
Percent Similarity:  43.42%
Best Local Similarity: 30.92%
Query Match:         3.16%
DB:                  13

                    Length:     820
                    Matches:     47
                    Conservative: 19
                    Mismatches:  61
                    Indels:      25
                    Gaps:         8
```

US-09-677-653A-50 (1-647) x BI650823 (1-820)

```
Qy   217 ArgIleArgValLeuArgProThrTyrAspValProAspProThrGluGly--------- 233
         ||:::||||                  :::           |||
Db    29 CGCCTGCGTACCTTGCAGCGACGTCCTCAGGTCCCCTCCCAAPCTGCCCGGGGTTCTCCTC 88

Qy   234 ----------LeuValArgThrValSerAspTyr------------ArgLeuThrTyr 245
         |||||||||||                  :::           |||||||||
Db    89 AGGTCACGTGCACGCGCGGTTCGGACCCTCGACAGCCATGTCCATCCTCTACGTCTCCTCA 148

Qy   246 LysAlaIleThrCysGluAlaAlaAsnMetPro----ThrLeuValAspGlyGlyPheTrpIle 264
         ||              |||  :::|||
Db   149 CCCGGACGCCTTCCCACGCCTCCAGCCCTCATCGCTGCCCGGGGAAGCTGGTG 208

Qy   265 ----------GlyGlyGlnTyrAlaLeuThrProThr------SerLeuProGlnTyr 278
         |||||||||||                  :::|||
Db   209 ACGGTCCCGGGTGGGGAG--GCCCTCACCCCCCATCTGCTTGCCAGCCTCCCCGA 262

Qy   279 AspValSerGluAlaTyrAlaLeuHisThrLeuThrPheAlaAlaArgProSerSerAlaAla 298
         |||              |||:::|
Db   263 ------------GCAGCCGGACTCCATTCCCTCCGCGCTGCGGGCGGGCCGGGCAGG 310

Qy   299 AlaLeuAlaPheValTrpAlaGlyLeuProGlnGlyGlyThrAlaProLeuAlaGlyThrPro 318
         ||||||                  :::           |||
Db   311 GACCTGGGCCGGGCCTGGCCGTCGGTGTGGGGGGCCCGGGCTGTAGCTCAGCTGTTGTGCCTGCAG 370
```

```
Qy    319 AlaTrpGluGlnAlaSerGlyGlyTyrLeuThrTrpArgHisAsnGlyThrThrPhe 338
          |||         |||         |||:::          |||||||||
Db    371 GCCTGGGGGACCAGCCGGGGGCAGCCGCCGGGGGCAGCTGTCCTGGTCCAGCAGCAGTCAGTTATG 430

Qy    339 ProAlaGlySerValSerThrTyrValLeuProGluGly 350
          |||    |||    |||||         |||
Db    431 CTGACACGGAGC---TCATACCCGCTGCCTGCGGGG 463
```

RESULT 14
CNS03M9N/c

LOCUS   CNS03M9N   959 bp   DNA   linear   GSS 17-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
            037K22 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL250484
VERSION   AL250484.1 GI:7971496
KEYWORDS   GSS; genome survey sequence.
SOURCE   Tetraodon nigroviridis
  ORGANISM Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 959)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE   Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 959)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE   Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 959)
  AUTHORS   Genoscope.
  TITLE   Direct Submission
  JOURNAL   Submitted (12-APR-2000)
  COMMENT   Genoscope.
            This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES   Location/Qualifiers
     source   1..959
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="037K22"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG037BF11LP1-end : T7"
BASE COUNT   167 a   284 c   332 g   175 t   1 others
ORIGIN

```
Qy    199 ValValAspSerGlyGlnTrpIleAsnPheThrAsnAspThrThrThrTyrTyrValArgIle 218
          |||:::         |||                                    |||
Db    835 ACCGTGAACGAGGAG---------------------------------------GTG 818

Qy    219 ArgValLeu-ArgProThrTyrAspValPro------AspProThrGluGly------- 233
          ::::|||||                |||                    |||:::||||
Db    817 AAGGTACGGAGCCCCGAGGGGAGACGCCCACGAGCGGTTCCTGGTCAGGGACGGGGCT 758

Qy    234 ------------------------------------------LeuValArgThrValSe 239
                                                    :::||::|||||::::
Db    757 GCAGCGCGCCTCTTCTCTTTGTCGTCACAGGCGCCTGACGGCGCCGAGATCGTGAAGACCATCAG 698

Qy    239 rAsp----------TyrArgLeuThrTyrLysAlaIleThrCysGluAlaAsnMetProTh 256
          :::                |||||::|||                    |:||
Db    697 AGACATCATCGCTCTCAACCCCTCTACAACAGTCAGTCCGCCGCCCCGGTCCACACCAC 638

Qy    256 rLeuValAspGlnGlyPheArgIleGlyGlyGlnTyrArgAlaLeuThrProThrSerLeuPr 276
          |||                    |||                    |||||
Db    637 GCTG---------------------------------------GCCTCCACACAGAAACGCCTG-- 613

Qy    276 oGlnTyrAspValSerGlyGluAlaTyrAlaLeuHisThrLeuThrProHisAlaArgProSerSe 296
                                                    |||
Db    612 -------------------------------------------CTGCCCGCTGTGGACGCACAGCCCCGGTCT 581

Qy    296 rAlaAlaAlaAlaLeuAlaAlaPheValThrProAlaGlyLeuProGlnGlyGlyThrAlaProAlaGl 316
          ::|||         |||||                |||::||||||||||
Db    580 CTGGTCTGCCTCAGACGCCGGCTCGATCAGACTCACCGGAGCTCGTGGTTCACCTGCTGG 521

Qy    316 yThrPro-----AlaTrpGluGlnAlaSerGlyGlyTyrLeuThrTrpArgHisAs 334
          |:::|||         ||||||         |||         |||
Db    520 GTCACCTGAGTCGGGCCTGGAGCCCACGGTCGGGTGTGGAACCTTAGCGTGGACGGAGAC 461

Qy    334 nGlyThrThrThrPhePheProAlaGlySerValSerThrTyrValLeuProGluGlyPheAsnValGl 354
          |||||         |||:::|||||::|||::||
Db    460 CGGACACATGTCCACCTCTGGCCCCAGGAGTCAGGCG------TCCAACTGTGTCT 410

Qy    354 uArgTyrArgAspProAsnAsp------------------------GlySerTrpTh 364
          :::                |||                         |||::||||
Db    409 CCTGTGGTCTCCAGGGAGGTCGGTCCTCCAGATGATCGCAGGCTGGCTTCAGAGGGGTGGGAC 350

Qy    364 rAspPheAlaSerAlaGlyAspThrValThrPheArgGlnValAlaValAlaAspGluVa 384
          |||:::                |||:::         |||||||:::
Db    349 A-------------ACCCCATCTACCTCACCGGCTCAGCCCTCAGACGGCACCGGACATGGGAGCAGCTCGACCCGAGCAGAGTCC 299

Qy    384 lValThrAsnAsnProAla-GlyLeuGlyGlySerAlaProThrProPheThrValArgValProP 404
          |||||         :::|||         ||||||:::||
Db    298 CACGAGCCTCCAGGACATCCTCGAGAGGCCCACGTGAGTCGGCGTGTGCCCCGGTCCTC 239

Qy    404 roSerAsnAlaThrTyrThrAsnThrValPheArgGlnThrLeuLeuGluThrArgProSerS 424
          |||                    |||         |||
Db    238 CGCCCAGG-----------------------------CTGGTTCGGTCGTCCCCTGATG 212

Qy    424 erArgArgLeuGluLeuProMetProProAlaAspPheGlyGlnThrValAlaAlaAsnAsp 444
          |||:::|||::         |||||::|||
Db    211 GGACCCGTGTCCTCCTGCAGATCCCCAAGGGCCTACAAGGCCCGTCCCTGCGTCCCTGAAGA 152

Qy    444 roLysIleGluGlnSerLeuLeuLysGlyThrLeuGlyCysTyrLeuValHisSerLysM 464
          :::                |||:::|||:||
Db    151 AGGAGTACGAGGCTGAGCAAGCTGCACGGCGCCCTGGGCCCCAGGAGGCTCAGACATGG 92

Qy    464 etArgAsnPro 467
          |||||
Db    :91 CGTCCCAGCCC 81
```

RESULT 15
BQ649851

LOCUS   BQ649851   983 bp   mRNA   linear   EST 15-JUL-2002
DEFINITION AGENCOURT_8341749 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6267739
            5', mRNA sequence.
ACCESSION   BQ649851
VERSION   BQ649851.1 GI:21774023
KEYWORDS   EST.

```
Qy    159 SerPheProMetPheArgThrAlaTyrValAlaValAlaAsnValGluAsnLysGluMet 178
          ::||||         |||||:::
Db    955 ACCTTTGACCTCCGCTCTTCTCCCCAGGTTTCAGAAGCCGACTTGAGCTTCCAAGCTTCAG 896

Qy    179 SerLeuAspValValAsnAspLeuIleGluTrpLeuAsnLeuAlaAspThrProArgTyr 198
          ||||||         |||         |||:::
Db    895 TCTTTGCCATGCTGCCCAACGTCTCTGATGGTGGAGGTGGAGAACATCCACCACGGCAGTTC 836
```

US-09-677-653A-50 (1-647) x CNS03M9N (1-959)

Alignment Scores:
Pred. No.:   2.75
Score: 106.00   Length: 959
Percent Similarity: 34.49%   Matches: 71
Best Local Similarity: 20.58%   Conservative: 48
Query Match: 3.14%   Mismatches: 138
DB: 17   Indels: 88
                                   Gaps: 11

```
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 983)
AUTHORS       NIH-MGC http://mgc.ncl.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              http://image.llnl.gov
              Plate: LLCM2438    row: 1    column: 20
              Tissue Procurement: CGAP (Stanford)
              cDNA Library Preparation: Rubin Laboratory
              DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              High quality sequence stop: 622.
EATURES       Location/Qualifiers
source        1..983
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:6267739"
              /clone_lib="NIH_MGC_100"
              /tissue_type="hepatocellular carcinoma, cell line"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:
              EcoRI; cDNA made by oligo-dT priming. Directionally cloned
              into EcoRI/XhoI sites using the following 5' adaptor:
              GGCACGAG(G). Size-selected >500bp for average insert size
              1.8kb. Library constructed by Ling Hong in the laboratory
              of Gerald M. Rubin (University of California, Berkeley)
              using ZAP-cDNA synthesis kit (Stratagene) and Superscript
              II RT (Life Technologies). Note: this is a NIH_MGC
              Library."

BASE COUNT    219 a   291 c   304 g   169 t
ORIGIN
```

```
US-09-677-653A-50 (1-647) x BQ649851 (1-983)

Alignment Scores:
Pred. NO.:             2.86
Score:               106.00                 Length:          983
Percent Similarity:   33.41%                Matches:          93
Best Local Similarity: 22.20%               Conservative:     47
Query Match:          3.14%                 Mismatches:      126
DB:                   14                    Indels:          154
                                            Gaps:             21
```

```
Qy    29 ValAsnGlyArgArgAsnGlnArgArgArgThrGlyArgGlnValSerProProAsn 48
         ::::::::::|||  |||  ||| |||||||  ||||||||
Db    10 CTGAGCGGTCGCAGCGCGCAGCCGCCGAGCGCCCGGCGGCCGCTGCTACTCCG------- 60

Qy    49 PheThrAlaAlaAlaGlnAspLeuAlaGlnSerLeuAspAlaAsnThrValThrPhePro 68
         ::::||                :::|||
Db    61 --------------------GAATCTGCTAATCCCAGTCGGGTGCCGATCCCC 93

Qy    69 AlaAsnIleSerSerMetProGluPheArgAsnTrp------------AlaLysGly 83
                       |||                    |||  :::
Db    94 AGCCCGCGCCCCATGGCCGCCT-----ACAAACTGGTGCTGATCCGGCACGGAGAGCG 147

Qy    84 LysIleAspLeuAspSerAspSerIleGlyTrpTyrPheLysTyrLeuAspProAlaGly 103
         :::::::||::::::||| |||              |||  ||||||||||
Db   148 CATGG--AACCTGGAGAACCCTTCAGCGGTCGTACGACCGCGACCTGAGCCCGGCGGGC 206

Qy   104 AlaThrGlySerAlaAlaArgAlaValGlyGluTyrSerLysIleProLeuProValLys 123
         ||:||::||  |||::||                 |||
Db   207 CACGAGGAGCGCGAACCGCGC---GGGCAG------GCGCTACGAGATGCTGGCTATGAG 257

Qy   124 PheSerValAspAlaGluIleIleArgGluIleTyrAsnGluGluCysProValValThrAsp 143
         |||                                            |||
Db   258 TTTGACATC----------------------------TGCTTCACCTCA 278
```

```
Qy   144 ValSerValProLeuAspGlyArgGlnTrpSerLeuSerIlePheSerPheProMetPhe 163
         |||                                       |||:::::
Db   279 GTGCGAGAAGAGAGGCGAGATCCGGACCCCTGGACAGTG------------------ 314

Qy   164 ArgThrAlaTyrValAlaValAlaAlaAsnValGluAsnLysGluMetSerLeuAspValVal 183
         |||                              |||::          |||||
Db   315 ------------------------------------CTAGATGCCATT 326

Qy   184 AsnAspLeuIleGluTrpLeuAsnAsnLeuAlaAspThrProArgThrTyrValValAspSerGlu 203
         |||           :::|||                    |||||
Db   327 GATCAGATG------TGGCTGCCAGTGGTGAGGACTTGGCGC--------------- 362

Qy   204 GlnTrpIleAsnPheThrAsnAspThrThrThrTyr----------- 214
         |||:::
Db   363 ----------CTCAATGAGGCGGCACTATGGGGGGTCTAACCGGTTGTCAATAAAGCA 407

Qy   215 -----------TyrValArgIleArgValLeuArgProThrThrTyrAspTyrVal 227
         :::::::::::|||           |||  :::|||||||||||
Db   408 GAAACTGCTGCAAAGCATGGTGAGGCCCAGTGAAGATCTGAAGCGGGGCGTCCTATGAGATGTC 467

Qy   228 ProAspProThrGluGlyLeuValArgThrValSerAspTyrArgLeuThrTyrLysAla 247
         |||                                          |||
Db   468 CCACCACCT----------------------------------------------- 476

Qy   248 IleThrCysGluAlaAsnMetProThrLeuValAlaAspGlnGlyPheTrpIleGly----- 265
         |||                              |||:::
Db   477 --------CCGATGGAGCGCCGACCATCCTTTCACAGCAACACATCAGT 515

Qy   266 ------GlyGlnTyrAla---LeuThrProThrSerLeuProGlnTyrAspValSerGlu 282
         |||           |||||       |||  |||||
Db   516 AAGGATCGCCAGGTTATGCAGACCTCACAGAAGATCCAGGTACCCCGTGGAG--------- 566

Qy   283 AlaThrTyrAlaLeuHisThrLeuThrProAlaArgProSerSerAlaAlaAlaLeuAlaAlaPhe 302
         |||           |||      |||  |||||||||||
Db   567 ------------AGTCTGAAGGATACTATATTCCCAGAGCTCTGCCCCTTC 602

Qy   303 --------ValTrpProLeuProGlnGlyGlyThrAlaProAlaGlyThrPro 318
         :::||| :::|||
Db   603 TGGAAATGAAGAAATAAGTTCCCCAGATCAAGGAGAAGGGAAAACTGTACTGATTGCAGCCA 662

Qy   319 AlaTrpGluGln-------- 333
         |||||                :::
Db   663 --TGGCAACAGCCTCCGGGCATTGTCAAGCATCCTGGGGGGAATAAAAGACGTGGCCCA 719

Qy   334 AsnGlyThrThrPhePheArgGlnValAlaValAspGluValValThrAsnAsnProAla 353
         :::||:|||||     |||
Db   720 CATGGGACTGAC--CCTGCCGACTGGGATTCCCATTGGCTATGGATTGGACCAAAACTG 776

Qy   354 GluArgThrTyrAsnProAsnAspGlySerTrpThrAsp------------PheAlaSerAlaGly 370
         |||  |||           |||||||||||              :::
Db   777 GAAAGCTATCAAGCCCATGCACAGGTTCCTGGGGGGATAAAAAGACGTGGCCCAAAGCCCTG 836

Qy   371 AspThrValThrPheArgGlnValAlaValAspGluValValThrAsnAsnProAla 390
         :::||||||     |||                  |||
Db   837 GAAACTGTGGGCTGGGCCCCAGGGCA--------------------AGGCCCAAA 869

Qy   391 GlyGlyGlySerAlaProThrPheThrVal---ArgValProProSerAsnAla 407
         |||  |||||       |||||||||||
Db   870 AATGGAAGGGCCCGCGGGGAAGGATACTGTCCCCGGGAAGCCCCTCCCTGCC 924
```

```
Search completed: February 27, 2003, 06:39:07
Job time : 1718 secs
```